PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-255172
(43)Date of publication of application : 01.10.1996

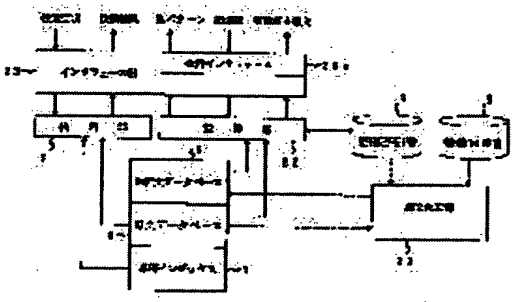
(51)Int.Cl. G06F 17/30

(21)Application number : 07-083458
(22)Date of filing : 16.03.1995

(71)Applicant : TOSHIBA CORP
(72)Inventor : SAKAI TETSUYA
MIIKE SEIJI
SUMITA KAZUO

(54) DOCUMENT RETRIEVAL SYSTEM

(57)Abstract:
PURPOSE: To lighten the burden of document retrieving operation on a user by extracting only desired extracted sentences or information from the whole text of a retrieved document and displaying it.
CONSTITUTION: The document retrieval system is equipped with an original text processing part 20 which sets plural kinds of sentence types for discriminating the contents of documents like an opinion and a proposal, generates extracted sentence data in sentence units classified by the types of the respective sentences from an original text database 6 storing original text data constituting the document, and stores the data as an extracted sentence database 5. The original text processing part 20 is equipped with a style decision part which extracts extracted sentence data corresponding to a specified sentence style and a shaping part which shapes the extracted sentence data into a specific style having, for example, conjunctions removed.



LEGAL STATUS

[Date of request for examination]	16.04.1999
[Date of sending the examiner's decision of rejection]	28.03.2000
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	
[Date of registration]	
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－255172

(43)公開日　平成８年(1996)10月１日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ　17/30		9194－5L	Ｇ０６Ｆ	15/401	３２０Ａ
		9194－5L		15/403	３８０Ｄ

審査請求　未請求　請求項の数６　ＦＤ　（全 23 頁）

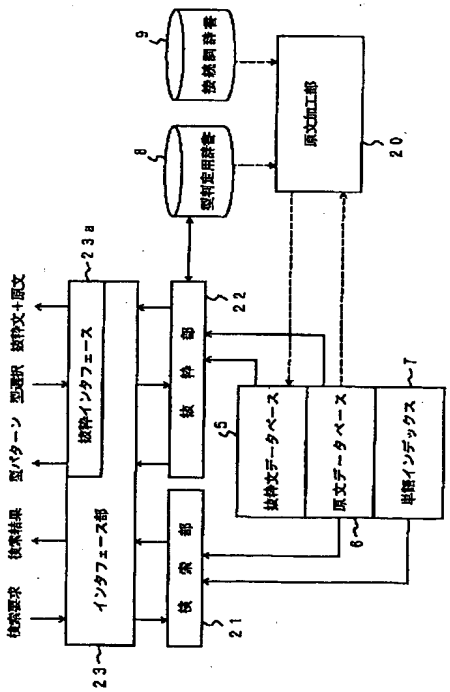
(21)出願番号　特願平7－83458

(22)出願日　平成７年(1995)３月16日

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者　酒井　哲也
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内
(72)発明者　三池　誠司
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内
(72)発明者　住田　一男
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内
(74)代理人　弁理士　鈴江　武彦

(54)【発明の名称】　文書検索システム

(57)【要約】　（修正有）
【目的】検索された文書の全文から所望の抜粋文や情報のみを抽出して表示できるようにして、利用者の文書検索作業に要する負荷の軽減化を図ることにある。
【構成】文書を構成する原文データを格納した原文データベース６から、例えば意見、提言等のように文章の内容を識別するための複数種類の文の型を設定し、この各文の型に分類した文単位の抜粋文データを作成し、抜粋文データベース５として格納する原文加工部２０を備えた文書検索システムである。原文加工部２０は、指定された文の型に対応する抜粋文データを抽出する型判定部と、抜粋文データを例えば接続詞を除去したような所定の形式に整形する整形部とを備える。



1

【 特許請求の範囲】
【 請求項1 】　検索要求に応じた文書を検索する文書検索システムにおいて、
検索対象の文書を構成する原文データを格納した原文データベース手段と、
前記原文データから予め設定した複数種の文の型に分類した文単位の抜粋文データを格納した抜粋文データベース手段と、
前記複数種の文の型から所望の文の型を選択する選択手段と、
この選択手段により選択された文の型に対応する抜粋文データを前記抜粋文データベース手段から抽出する抽出手段と、
この抽出手段により抽出された全ての抜粋文データを一覧的に表示する表示手段とを具備したことを特徴とする文書検索システム。
【 請求項2 】　検索要求に応じた文書を検索する文書検索システムにおいて、
検索対象の文書を構成する原文データを格納した原文データベース手段と、
前記原文データから予め設定した複数種の文の型に分類した文単位の抜粋文データを作成する作成手段と、
この作成手段により作成された各抜粋文データを所定の形式に整形する整形手段と、
この整形手段により整形された各抜粋文データを格納した抜粋文データベース手段と、
前記複数種の文の型から所望の文の型を選択する選択手段と、
この選択手段により選択された文の型に対応する抜粋文データを前記抜粋文データベース手段から抽出する抽出手段と、
この抽出手段により抽出された全ての抜粋文データを一覧的に表示し、かつ抽出された前記各抜粋文データに対応する前記原文データを強調的に表示する表示手段とを具備したことを特徴とする文書検索システム。
【 請求項3 】　検索要求に応じた文書を検索する文書検索システムにおいて、
検索対象の文書を構成する原文データを格納した原文データベース手段と、
前記原文データから予め設定した複数種の文の型に分類した文単位の抜粋文データを作成する作成手段と、
この作成手段により作成された各抜粋文データを所定の形式に整形する整形手段と、
この整形手段により整形された各抜粋文データを格納した抜粋文データベース手段と、
前記複数種の文の型から所望の文の型または全文表示モードを選択する選択手段と、
この選択手段により選択された文の型に対応する抜粋文データを前記抜粋文データベース手段から抽出する抽出手段と、

2

前記選択手段により前記全文表示モードを選択された場合には、前記原文データベース手段から検索対象の文書の全文に対応する原文データを取出して出力する全文出力手段と、
前記抽出手段により抽出された全ての抜粋文データまたは前記全文出力手段により出力された原文データを表示する表示手段とを具備したことを特徴とする文書検索システム。
【 請求項4 】　検索要求に応じた文書を検索する文書検索システムにおいて、
検索対象の文書を構成する原文データを格納した原文データベース手段と、
前記原文データから予め設定した複数種の文の型に分類した文単位の抜粋文データを作成する作成手段と、
この作成手段により作成された各抜粋文データを所定の形式に整形する整形手段と、
この整形手段により整形された各抜粋文データを格納した抜粋文データベース手段と、
前記複数種の文の型から所望の文の型を選択する選択手段と、
この選択手段により選択された文の型に対応する抜粋文データを前記抜粋文データベース手段から抽出する抽出手段と、
予め指定したキーワードに基づいて、前記抽出手段が抽出された抜粋文データの中で前記キーワードを含む特定の抜粋文データのみを選択して表示する表示手段とを具備したことを特徴とする文書検索システム。
【 請求項5 】　検索要求に応じた文書を検索する文書検索システムにおいて、
検索対象の文書を構成する原文データを格納した原文データベース手段と、
前記原文データに対して文単位の言語解析処理を実行し、この言語解析処理により得られた言語情報を生成する言語情報生成手段と、
前記原文データから予め設定した複数種の文の型に分類した文単位の抜粋文データであって、前記言語情報を付加した抜粋文データを格納した抜粋文データベース手段と、
前記複数種の文の型から所望の文の型を選択する選択手段と、
この選択手段により選択された文の型に対応する抜粋文データを前記抜粋文データベース手段から抽出する抽出手段と、
この抽出手段により抽出された全ての抜粋文データを前記言語情報を伴って一覧的に表示する表示手段とを具備したことを特徴とする文書検索システム。
【 請求項6 】　検索要求に応じた文書を検索する文書検索システムにおいて、
検索対象の文書を構成する原文データを格納した原文データベース手段と、

前記原文データに対して言語解析処理を実行し、この言語解析処理により得られた言語情報に基づいた文の所定の構成要素を前記原文データから抽出する言語解析処理手段と、

この言語解析処理手段により抽出された前記所定の構成要素毎に予め設定した複数種の情報の型に分類した抜粋情報データを格納した抜粋情報データベース手段と、

前記複数種の情報の型から所望の情報の型を選択する選択手段と、

この選択手段により選択された情報の型に対応する抜粋情報データを前記抜粋情報データベース手段から抽出する抽出手段と、

この抽出手段により抽出された全ての抜粋情報データを一覧的に表示する表示手段とを具備したことを特徴とする文書検索システム。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、特に文書単位に文書情報を検索するシステムにおいて、文書の原文を加工処理して得られた抜粋文または抜粋情報を検索する機能を備えた文書検索システムに関する。

【０００２】

【従来の技術】従来、フルテキストサーチ方式の文書検索システムは、文書データベースから必要な文書や情報を検索する場合に、文書を構成する文字または単語の全てを検索対象とするため、例えばキーワード検索方式と比較して、キーワードを付与する作業を必要せず、また利用者がそのキーワードに精通している必要がないなどの利点を備えている。

【０００３】しかしながら、フルテキストサーチ方式では入力した単語や文章等の検索要求に対して、その検索要求に合致する文書数と文書データ量は膨大である。この方式では、検索処理により検索要求に合致する各文書の全文が、そのままの状態で表示画面上に表示される。利用者は、表示画面上で検索された各文書の全文を確認し、この中から必要な文書または情報（文章や単語からなる）を取出すことになる。

【０００４】したがって、利用者は、検索された膨大な量の文書の原文に目を通す必要があるため、必要な文書や情報を得るまでに多大な時間と労力が必要となる。このような問題点を解消するためには、検索された文書の原文から必要な文や情報のみを抽出して、見易い形式で表示画面上に表示するシステムが望ましい。

【０００５】近年、文書から自動的に抄録や要約を作成する技術を開発されており、この技術を利用して検索された文書中から重要文のみを抽出して表示するシステムが実現されている。また、文書中から例えば結論に相当する章または項の文章を抽出して表示するシステムも開発されている。

【０００６】このようなシステムにおいても、同一文書

の中でどれが重要文または重要情報であるかは、観点の相違や個人差に左右されるので、必ずしも利用者にとって適切な文や情報が抽出されるとは限らない。また、例えば結論に相当する章または項の文章が抽出されても、その中に必要な文や情報が含まれているとは限らない。さらに、重要文のみを抄録として表示するシステムでは、接続詞などによる文間の関係が壊され、誤解を生むような文等が表示される可能性がある。

【０００７】

【発明が解決しようとする課題】前述したように、従来の文書検索システムでは、膨大な量の検索結果から必要な文書や情報を取出すため、利用者には多大な時間と労力が必要となり、負担が大きい。また、文書中から抄録や要約を作成したり、指定の章または項の部分を抽出する技術を利用して、検索された文書から抜粋部分を表示するシステムが開発されている。しかし、表示された抜粋部分が、利用者にとって必ずしも適切であるとは限らず、結果的に利用者の必要な抜粋文や情報が得られない場合も多い。

【０００８】本発明の第1の目的は、検索された文書の全文から所望の抜粋文や情報のみを抽出して表示できるようにして、利用者の文書検索作業に要する負荷の軽減化を図ることにある。

【０００９】本発明の第2の目的は、同一文書に対して様々な観点や基準に基づいて抜粋文や情報を抽出して表示できるようにして、多様な検索目的に応じた検索処理を実現することにある。

【００１０】本発明の第3の目的は、検索された文書の全文から所定の基準に基づいて抜粋文や情報のみを抽出して表示する場合に、優先度の順序に従って表示する抜粋文や情報の量を調整できる検索処理を実現することにある。

【００１１】

【課題を解決するための手段】本発明は、文書を構成する原文データを格納した原文データベースから、例えば意見、提言等のように文章の内容を識別するための複数種類の文の型を設定し、この各文の型に分類した文単位の抜粋文データを作成し、抜粋文データベースとして格納する手段を備えた文書検索システムである。さらに、本システムは、抜粋文データを例えば接続詞を除去したような所定の形式に整形する整形手段と、指定された文の型に対応する抜粋文データを抽出する抽出手段と、抽出された全ての抜粋文データを一覧的に表示する表示手段とを備えている。

【００１２】

【作用】本発明では、利用者が予め設定されている複数の文の型から所望の型を選択すると、抽出手段は抜粋文データベースから文の型に対応する抜粋文データを抽出する。例えば意見、提言の文の型を選択すると、検索された文書からその型に分類された全ての抜粋文データが

5

抽出される。表示手段は、抽出された全ての抜粋文データを一覧的に表示する。これにより、利用者は、表示画面上において、例えば意見、提言の文の型に含まれる抜粋文データのみを確認するだけで、必要な文章や情報を得ることができる。また、利用者は文の型を変更するだけで、変更した文の型に含まれる抜粋文データを表示画面上で得ることができる。したがって、文の型を指定するだけで、利用者の必要な文、情報、およびそれらを含む文書を簡単に検索することができることになる。

【００１３】

【実施例】以下図面を参照して本発明の実施例を説明する。

（システムの構成）図1は本実施例に係わる文書検索システムの要部を示すブロック図である。

【００１４】本システムは概略的に、本実施例の文書検索処理に関係する各種処理及びシステム全体の制御を行なう中央処理部（ＣＰＵ）１、表示コントローラ２、表示部3、記憶部4、入力コントローラ１０および入力部１１からなる。

【００１５】表示部3は例えば液晶ディスプレイからなり、表示コントローラ２の制御により検索結果や文書等を表示する表示画面を有する。入力部１１は入力コントローラ１０の制御により、検索命令や各種の選択、指定等の入力を行なうキーボード１１ａとマウス１１ｂを有する。

【００１６】記憶部4は、ＣＰＵ1の各種処理を行なうためのメインメモリ（ＲＡＭ）、及びハードディスク装置等のファイル装置からなり、後述する本実施例に関係する抜粋文データベース５、原文データベース６、単語インデックス７、型判定用辞書8及び接続詞辞書9を格納している。

（第1の実施例）前記のシステムにおいて、第1の実施例として図2に示すような機能ブロック図を示す。図2において、原文加工部２０、検索部２１、抜粋部２２、インターフェース部２３および抜粋インターフェース２３ａは、ＣＰＵ1が実行するプログラムに対応している。

【００１７】インターフェース部２３は、利用者の検索要求をキーワード、あるいは自然言語からなる文または文章の形式で受付け、これを検索部２１に渡す。検索部２１は、原文データベース６及び単語インデックス７を参照し、該当する文書のリストやその内容を検索結果としてインタフェース部２３に渡す。インタフェース部２３はその検索結果を表示部3の表示画面上で、利用者に提示する。

【００１８】インターフェース部２３に組込まれている抜粋インターフェース２３ａは、検索結果として提示される個々の文書の全文の中から、利用者の指定した文の型に対応する文のみを抜粋して表示画面上に表示する機能を有する。

6

【００１９】利用者が文の型を選択すると、抜粋インタフェース２３ａは選択された型を抜粋部２２に渡す。抜粋部２２は、原文データベース６及び抜粋文データベース５を参照し、全文の中から該当する型の文のみを抜粋し、文中の接続詞を除去して抜粋インタフェース２３ａに渡す。抜粋インタフェース２３ａはこれらの抜粋文を原文と対比させながら表示画面上に表示する。

【００２０】原文加工部２０は、図3に示すように、大別して型判定部２０ａおよび整形部２０ｂからなり、原文データベース６から抜粋文データベース５を作成する。

【００２１】型判定部２０ａは、型判定用辞書8を参照し、与えられた文書の各文がどの型に当てはまるかを判定する。換言すれば、予め設定された複数の型に、文書の各文を分類するための判定処理を実行する。一方、整形部２０ｂは、接続詞辞書9を参照し、与えられた文書の各文を所定の形式、即ち本実施例では接続詞を除去した形式に整形処理する。

【００２２】原文加工部２０は、利用者が型の選択をする度にリアルタイムに実行して、抜粋文データベース５を作成する方式でもよいし、またはシステムの使用前（検索処理前）にバッチ処理的に抜粋文データベース５を作成する方式でもよい。本実施例では、高速処理の可能な後者の方式を想定して説明する。

【００２３】なお、図2において、破線で描かれた矢印はシステム使用前のデータの流れを意味し、実線で書かれた矢印はシステム使用時のデータの流れを意味する。

（原文加工処理）前記のような原文加工部２０の具体的動作を図3乃至図5を参照して説明する。

【００２４】まず、原文加工部２０は、原文データベース６から文書単位に文書を構成する各文（原データ）を読み込む（ステップＳ１）。原文データは、図3に示すように、例えば「日本的な雇用は、…と言われている。」や「その背景には…と言う事実がある。」等の各文である。

【００２５】次に、型判定部２０ａは型判定用辞書8を参照し、各文に該当する型を判定し、その型を示すタグを付与する（ステップＳ２）。ここで、型判定用辞書8は、例えば図5（Ａ）に示すように、予め設定された文の型とその型に対応する複数のパターンとを組とする各項目からなる。パターンは、例えばその文の型を的確に表現する文節からなる。

【００２６】具体例として、図5（Ａ）に示すように、例えば文の型として「意見・提言」、「問題提起」、「「予想・推定」等が設定されている。「意見・提言」の型には、例えば「…だと考えられる」や「…べきではないだろうか」等のパターンがある。この具体例では、パターン「…であると思う」は「意見・提言」の型と「予想・推定」の型の両方に割り当てられている。即ち、このパターンを含む文は、「意見・提言」型の文を

表現するし、また「 予想・ 推定」型の文も表現する。このように、複数の型に合致する文もあるため、各型は必ずしも相互に排他的ではない。

【００２７】また、型判定用辞書8 に用意する型の種類は、例えば取り扱う 文書の種類に応じて設定する方式でもよい。例えば新聞記事の検索処理では「 事実」、「 発言」等の型を用意し、技術論文の検索処理では「 方法」や「 結論」などの型を用意する。

【００２８】さらに、整形部２０ｂは接続詞辞書9 を参照し、各文を所定の形式、本実施例では接続詞を除去した形式に整形処理する（ ステップＳ３）。具体例として、図3 に示すように、例えば「 ところが、企業は…変化した。」と言う 文から接続司「 ところが、」を除去する。

【００２９】原文加工部２０は、前記のように加工処理した抜粋文データを抜粋文データベース５に蓄積する（ ステップＳ４）。即ち、原文加工部２０は、図3 に示すように、原文データ６から判定された型に分類し、接続詞を除去した抜粋文データ５ａを作成する。例えば「 ところが、企業は…変化した。」という 原文は、「 …した。」というパターンにマッチする「 過去」という 型のタグが付与され、さらに接続詞「 ところが、」を除去された抜粋文に変換される。

【００３０】本実施例では、利用者は、後述するように、検索結果の個々の文書からどのような型の文を抜粋して表示するかを指定することができる。この場合、全文の中から「 意見・ 提言」の型の文のみを抜粋表示したり、「 結論」の型の文のみを抜粋表示したりしたりして、型を切替えることが可能である。また、図5（ Ｂ）に示すように、例えば「 意見・ 提言」と「 予想・ 推定」という 二つの型を同時に選択してもよい。前述したように、文書の全文において、複数の型に関係する文が存在するため、図中のベン図のように、斜線部分の範囲の文が抜粋されるとになる。なお、全文の中で、設定した型に含まれない文も当然ながら存在する。

（ 抜粋処理）次に、抜粋インタフェース２３ａによる抜粋文の表示処理について、図6 と図7 を参照して説明する。

【００３１】まず、利用者は、図6（ Ａ）に示すように、表示部3 の表示画面上に表示された抜粋文の型を選択するためのメニュー画面から、所望の型を入力部１１のマウス１１ｂを操作して選択する。ここで、抜粋インタフェース２３ａは、抜粋部２２を介して抜粋文データベース５から文書に対応する抜粋文データを取り出す（ ステップＳ１０）。

【００３２】抜粋インタフェース２３ａは、抜粋文データから両者が選択した型のタグが付与された抜粋文の全てを抽出する（ ステップＳ１１）。この抽出した抜粋文の文番号を記憶する（ ステップＳ１２）。そして、表示画面上に抜粋画面に例えば箇条書き形式で、抽出した全

ての抜粋文を表示する（ ステップＳ１３）。

【００３３】具体例として、図6（ Ａ）に示すように、型として「 意見・ 提言」が選択された場合に、その型に合致する抜粋文（ ｂ，ｅ，ｈ，ｉ）を表示する。

【００３４】ここで、表示画面上において、抜粋インタフェース２３ａは、抜粋画面と原文画面を並べて表示する。抜粋インタフェース２３ａは、抜粋部２２を介して原文データベース６から該当する文書の原文データを取り出す（ ステップＳ１４）。そして、原文画面上に取り出した原文データを表示する。このとき、抜粋インタフェース２３ａは、記憶した抜粋文の文番号に対応する原文（ ｂ，ｅ，ｈ，ｉ）を例えばカラー表示やアンダーライン表示等による強調表示する（ ステップＳ１５）。

【００３５】即ち、原文画面上において、「 意見・ 提言」型に合致して、接続詞「 しかし、」を除去した抜粋文b に対応する文番号2 の原文が強調表示される。また、抜粋文画面上に抜粋文i が表示されているが、これは原文画面中に収まりきらなかった文のうち「 意見・ 提言」の型に該当した文である。

【００３６】ここで、同一文書において、図6（ Ｂ）に示すように、利用者が型を「 予想・ 推定」に切替えたとき、原文画面中の文番号4，５の原文は「 予想・ 推定」の型に該当するため抜粋されている。この場合、文番号５に対応する抜粋文データｅは、「 意見・ 提言」の型と「 予想・ 推定」の型の両方に該当している。

（ パターン表示処理）以上のように、利用者が抜粋文の型を選択するには、選択肢として表示されている各型の意味を利用者にわかりやすく表示することが望ましい。例えば、「 意見・ 提言」という 型を選択すると、実際には「 …だと考えられる」や「 …べきではないだろうか」のようなパターンの文が抜粋されるということを利用者が知っていれば、所望の情報を得やすいと考えられる。

【００３７】そこで、図8（ Ａ）に示すように、利用者が型を選択すると、抜粋部２２が型判定用辞書8 から選択した型に対応する全てのパターンを取出して抜粋インタフェース２３ａに出力する（ ステップＳ２０，Ｓ２１）。抜粋インタフェース２３ａは、同図（ Ｂ）に示すように、例えば選択した「 意見・ 提言」の型に対応するパターンを列挙した一覧表示を行なう（ ステップＳ２２）。

【００３８】この機能により、利用者は、どの型を選択すると、どのような表現を含む文が抜粋されるかを予め知ることができる。

（ 選択表示処理）これまでの説明では、抜粋文データにおいて選択中の型のタグがつけられている文はすべて抜粋文として表示するものとしていた。この変形例として、選択中の型のタグがついている文をとりだした後、この中からある尺度に基づいて一つ以上の重要文を選択し、それを表示する方式を説明する。

【００３９】まず、図9 に示すように、選択した型のタ

グが付与された抜粋文の全てを抽出するまでの処理は、これまでの処理と同様である（ステップＳ３０，Ｓ３１）。

【００４０】次に、検索要求に含まれる検索キーワードを最も多く含む文が重要文であるとし、この尺度に基づいて該当する抜粋文を抜粋画面上に箇条書き形式で表示する（ステップＳ３２，Ｓ３３）。

【００４１】具体例として、図１０に示すように、選択した例えば型「目的・ねらい」に対して、全文6 文の中で、該当する文である文番号2，４の各文が抜粋文として抽出される。この2 文の中で、文番号2 の文は「分散型」と「アーキテクチャ」という２つの検索キーワードを含むが、文番号4 の文はひとつも含んでいない。したがって、文番号2 の文の方が文番号4 よりも重要な文であるとし、この文番号2 の文のみを抜粋文として表示する。

【００４２】これにより、利用者は、選択した型と共に、検索要求の検索キーワードを含む最も適切な抜粋文を参照できる可能性が高くなる。

（特定文表示処理）これまでの説明では、利用者は型の選択権のみをもち、ある型が選択された場合、その型に対応する全てのパターンにマッチする文が抜粋されることになっていた。この変形例として、特定のパターンにマッチする文のみを抜粋するように利用者に指定させる方式を説明する。

【００４３】まず、図１１に示すように、選択した型に対応するパターンを、型判定辞書8 から取り出すまでの処理は、これまでのパターン表示処理と同様である（ステップＳ４０，Ｓ４１）。

【００４４】ここでは、図１２（Ａ）に示すように、例えば「意見・提言」の型に該当するパターンを取り出している。このとき、抜粋文画面に各パターンと共に、各パターンにマッチする文の数を表示する（ステップＳ４２）。この例では、「…だと考えられる」のパターンの文は5 文、「…であると思う」のパターンの文は3 文であることが表示されているので、この二つのパターンを選択すると抜粋文画面には合計8 文が表示されることが事前にわかる。同様に、各パターンにマッチした文の長さや行数などの情報を表示すれば、利用者は画面にちょうど収まるくらいの分量の文を抜粋表示させることもできる。

【００４５】この表示されたパターンの中から、利用者が例えば「…だと考えられる」と「…であると思う」の各パターンを選択すると、抜粋部２２は抜粋文データベース5 から該当する抜粋文データを取出して抜粋インタフェース２３ａに返す（ステップＳ４４）。

【００４６】抜粋インタフェース２３ａは、図１２（Ａ）に示すように、利用者が選択した型に該当し、かつ選択した特定パターンにマッチする抜粋文のみを抜粋文画面に表示する（ステップＳ４５）。この場合、前述

のパターンにマッチした抜粋文の中から重要文のみを表示する処理を適用してもよい。

【００４７】以上のように、型の選択に加えてパターンの選択も利用者に行わせると、利用者は抜粋文の分量をある程度制御することができる。

（抜粋文の分量制御）次に、前記の変形例として、型判定辞書8 に用意したパターンに予め優先順位を設定し、パターンの優先順位に基づいて表示する文の分量を制御する方式について説明する。

【００４８】この変形例では、図１４（Ａ）に示すように、型判定辞書8 の構造として、型とパターンとの項目以外に、パターン間の優先順位を示す情報が含まれている。具体例として、「意見・提言」という型に対応するパターンの中では、「…だと考えられる」というパターンの優先順位が一番高いことが示されている。

【００４９】ここで、現在選択中の文書について、「意見・提言」の型に対応する各パターンにマッチする文の数をカウントすると、図１４（Ｂ）に示すような文数情報が得られると想定する。即ち、「…だと考えられる」と「…べきではないだろうか」のパターンの文のみ表示すれば合計3 文、さらに次に優先度の高い「…を提案する」のパターンの文まで表示すれば合計6 文が表示されることになる。

【００５０】この変形例では利用者が、図１３に示すように、表示画面上において抜粋表示する文の数（表示文数）ｎを指定する（ステップＳ５０）。ここでは、例えば図１４（Ｃ）に示すように、利用者が表示文数n を５文以内に制限したと想定する。

【００５１】抜粋部２２は優先順位が記述されている型判定辞書8 を参照し、利用者の指定した分量n と優先順位i までのパターンにマッチする文の分量mとを比較する（ステップＳ５１，Ｓ５２）。そして、優先順位の高いパターンにマッチする抜粋文を、指定分量n の範囲内で段階的に選択していく（ステップＳ５４～Ｓ５５）。最終的に、指定分量n の範囲内でかつ優先順位の高いパターンにマッチする抜粋文を抜粋インタフェース２３ａに出力する。

【００５２】抜粋インタフェース２３ａは、図１４（Ｃ）に示すように、利用者が指定した表示文数（５文）の範囲内の抜粋文（ここでは3 文）を抜粋画面に表示する（ステップＳ５６）。ここでは、優先順位が1 位と2 の「…だと考えられる」と「…べきではないだろうか」の各パターンにマッチする抜粋文のみを表示する。

【００５３】なお、ここでは文の分量の例として文数で説明したが、文の長さ、行数などをもとに表示の制御をすることも可能である。また、利用者に文の分量を指定させるかわりに、優先順位何番目までのパターンにマッチする文を表示させるかを指定させてもよい。

（型判定辞書の更新処理）これまでの説明では、型判定辞書8 は事前に用意された固定的なものを想定してい

た。この変形例では、型判定辞書8 に新たなパターンを追加登録したり、または逆に不適切と思われるパターンを削除する方式を説明する。

【００５４】まず、利用者の入力に応じて、型判定辞書8 の更新用画面を表示する（ステップＳ６０）。次に、更新対象の型とパターンを入力する（ステップＳ６１）。具体例としては、例えば図１７（Ａ）に示すように、利用者が原文画面上で、パターンとして登録したい範囲（ここでは、すべきだ）をマウス１１ｂにより指定する。そして、同図（Ｂ）に示すように、登録対象の型を指定する。ここでは、「意見・提言」の型に「…すべきだ」という表現のパターンを追加する場合である。

【００５５】これらの入力に応じて、抜粋部２２は型判定辞書8 の該当する型に関するエントリを検索する（ステップＳ６２）。そして、該当する型のエントリとして、入力した追加パターンを型判定辞書8 に登録して更新する（ステップＳ６３）。

【００５６】具体例として、図１６（Ｂ）に示すように、「意見・提言」の型のパターンのひとつとして、「…すべきだ」という表現を追加する。

【００５７】一方、追加だけでなく、逆に指定のパターンを削除する更新処理もある。具体例として、図１６（Ａ）に示すように、例えば、「意見・提言」の型に対応するエントリとして「…べきではないだろうか」という記述があった場合、利用者がこれを好ましくないと思う場合もあると考えられる。このような場合には、更新処理として削除を指定し、「…べきではないだろうか」という表現のパターンを削除する。

【００５８】このような機能により、利用者にとって、予め設定されている型判定辞書8 に不備があり、用意されている型に対応するパターンの中に必要と思われるパターンが登録されていない場合には、そのパターンを登録することができる。また、逆に型の表現としては不適切と思われるパターンが存在する場合には、そのパターンを削除することができる。型判定辞書の各エントリは、型とパターンの対という単純な構造をしているので、利用者がこれらの追加あるいは削除を行うことは容易である。したがって、利用者にとって有効な型判定辞書8 を再構築することが可能となる。

（優先度修正処理）前述の分量制御処理において、型判定辞書8 のパターン間に優先順位を付加する方式について説明したが（図１３と図１４を参照）、この優先順位を利用者に適応するように修正する方式について説明する。

【００５９】まず、図１９（Ｂ）に示すように、選択した型に対応する抜粋文データを抜粋画面上に表示する（ステップＳ７０）。ここで、抜粋文データを選択するための型判定辞書8 には、図１９（Ａ）に示すように、各パターンには優先順位を示す情報が付加されている。

【００６０】次に、利用者がマウス１１ｂにより、抜粋

画面上において重要だと思われる抜粋文を指定する（ステップＳ７１）。ここでは、図１９（Ｂ）に示すように、具体例として、「…だと考えられる」というパターンの文と「…すべきではないか」というパターンの文が表示されているが、利用者は、内容的にみて「…すべきではないか」の文が重要であると判断し、マウス１１ｂにより指定したと想定する。

【００６１】抜粋部２２は指定された抜粋文データから、その型に含まれるパターンを抽出し、型判定用辞書8 においてそのパターンの優先順位を高い方に修正する処理を行なう（ステップＳ７２，Ｓ７３）。

【００６２】具体例として、図１９（Ａ）に示すように、「意見・提言」の型のパターンとして、当初の型判定用辞書8 では、「…だと考えられる」のパターンの優先順位が第1 位で、「…べきではないだろうか」のパターンの優先順位が第2 位である。ここで利用者の「…べきではないだろうか」のパターンの優先度の修正を指定したことにより、図１９（Ａ）に示すように、両者の優先順位が逆転するように、型判定用辞書8 の内容が更新される。

【００６３】これにより、以後の分量制御処理を伴う抜粋処理において、更新した型判定用辞書8 を使用することにより、利用者に適応した抜粋文を抽出して表示することになる。

（複数文書の表示処理）これまでの説明では、検索結果の中の一文書を選択した時の処理が中心であった。この変形例では、検索結果の中の複数の文書に対して同一の型やパターンにより、抜粋処理を行なう方式について説明する。

【００６４】ここで、一般的に、検索結果は、検索要求に近いものから順に提示されることが多いが、選択した型やパターンに該当する抜粋文が多い順に文書を提示する方法も考えられる。このようにすれば、あとの方に提示される文書ほど抜粋される情報が少ないことがわかるので、複数の文書から効率的に情報を得ることが可能となる。

【００６５】まず、図２０に示すように、利用者の入力に応じて型とパターンが選択されると、検索されている複数の文書（ここではＡ，Ｂ，Ｃの3 文書）をソートする処理を行なう（ステップＳ８０～Ｓ８２）。そして、ソートした順序で各文書を表示する（ステップＳ８３）。

【００６６】ここで、図２１（Ａ）に示すように、検索された各文書Ａ，Ｂ，Ｃの抜粋文データには、各型やパターンに該当する文がいくつあったかという情報が記述されている。例えば、文書Ａの抜粋文データには、「意見・提言」の型に該当する文が１０文、「問題提起」の型に該当する文が0 文、「予想・推定」の型に該当する文が4 文含まれることなどが記されている。

【００６７】利用者の選択している型が「意見・提言」

である場合、この型に該当する文の数はA，B，Cの順で多いので、図２１（B）に示すように、この順序で文書のリストを表示する。利用者の選択している型が「予想・推定」である場合、この型に該当する文の数はB，C，Aの順で多いので、図２１（C）に示すように、この順序でリストを表示する。

（第2の実施例）次に、本発明の第2の実施例を、図２２に示す機能ブロック図を参照して説明する。本実施例は、前述した第1の実施例との相違点は、抜粋インタフェース２３ｂが原文表示専用の画面をもたないということである。以下、第1の実施例と異なる点について説明する。したがって、異なる点以外の機能は第1の実施例と同様である。

（全文選択処理）本実施例は、図２４に示すように、予め設定される複数の抜粋文の型の中に、「全文」という項目を含む。この項目を選択すると、抜粋画面上に、加工処理されていない原文の全文を直接表示する。

【００６８】即ち、図２３のフローチャートに示すように、両者が型を選択する場合に、「全文」の項目を選択すると、抜粋部２２は原文データベース５から原文データからなる全文を取り出し、抜粋インタフェース２３ｂに出力する（ステップＳ９０，Ｓ９１，Ｓ９５）。抜粋インタフェース２３ｂは、その原文データを抜粋文データとして抜粋画面上に表示する（ステップＳ９６）。

【００６９】一方、「全文」以外の型を選択した場合には、前述の第1の実施例と同様に、抜粋文データベース５から選択した型のタグが付与された抜粋文を全て取り出して、抜粋画面上に箇条書き形式で表示する（ステップＳ９１のNO，Ｓ９２～Ｓ９４）。具体例として、図２４に示すように、「意見・提言」の型が選択されると、その型に対応する抜粋画面上に表示する。

【００７０】本実施例によれば、原文画面を用意することなく、「全文」という型を用意するだけで、原文データを抜粋画面上で直接参照することが可能となる。したがって、利用者は、必要に応じて抜粋文データまたは原文データを簡単に選択して参照することが可能である。

（第3の実施例）次に、本発明の第3の実施例を、図２５に示す機能ブロック図を参照して説明する。本実施例の特徴は、文の型を判定するためのパターンとして、前述の第1と第2の実施例とは異なり、言語解析処理により得られる言語情報を利用する点にある。

【００７１】即ち、前述の第1と第2の実施例では、文の型を判定するためのパターンは、例えば「…だと考えられる」のような表層文字列である。即ち、「…だと考えられる」のような表層文字列を含む抜粋文は、例えば「意見・提言」の型に相当すると判定する。

【００７２】これに対して、本実施例は、原文の言語解析により得られる形態素情報（品詞、活用形等）や構文情報（主語、述語、係り受け等）の高度な言語情報を利用したパターンマッチング処理を実行して、最適な抜粋

文を表示する。

【００７３】以下、前述の第1と第2の実施例との相違点のみについて説明する。したがって、異なる点以外の機能は第1と第2の実施例と同様である。

（原文加工処理）図２６のフローチャートに示すように、まず、原文加工部２０は、原文データベース６から文書を構成する各文（原文データ）を読み込む（ステップＳ１００）。原文データは、図２７に示すように、例えば「日本的な雇用は、…である。」や「おそらく、その背景には…であろう。」等の各文である。

【００７４】ここで、図２７に示すように、原文加工部２０の解析部２０ｃは、予め用意された解析辞書２４を参照して言語解析処理を実行し、形態素情報や構文情報等の言語情報を抽出する（ステップＳ１０１）。

【００７５】型判定部２０ａは型判定用辞書8を参照し、各文に該当する型を判定し、その型を示すタグを付与する（ステップＳ１０２）。そして、接続司辞書9を参照した整形処理の後に、抜粋文データを抜粋文データベース５に格納する（ステップＳ１０３，Ｓ１０４）。このとき、抜粋文データには、型のタグと共に形態素情報、構文情報などの言語情報が付与されている。

【００７６】具体例として、図２７に示すように、例えば「日本的な雇用は、…である。」という抜粋文データには、「事実」の型を示すタグと形態素情報として「断定の助動詞」の言語情報が付与されている。

【００７７】このような言語情報を利用して、利用者が選択した型に対応するパターンマッチング処理を行なう場合に、型判定用辞書8には各型に対応するパターンとして前記の言語情報が用意されている。

【００７８】具体的には、図２８に示すように、パターンとして表層文字列を列挙する代わりに、例えば言語情報として品詞の種類に関する情報が設定されている。さらに、この言語情報と表層文字列を組合わせたパターンを用意するのが望ましい。

（第4の実施例）さらに、本発明の第4の実施例を、図２９に示す機能ブロック図を参照して説明する。本実施例の特徴は、前述の第1乃至第3の実施例とは異なり、型の判定処理と抜粋処理を文単位に限定せずに、文の様々な構成要素を単位とする点にある。即ち、句や節、特定の意味内容を表す表現など、文の様々な構成要素を抽出して表示する。

【００７９】本実施例では、抜粋処理により得られるデータを抜粋情報データと称し、原文加工部２０の加工処理により得られた抜粋情報データは抜粋情報データベース２５に格納される。

（原文加工処理）以下、図３１と図３２を参照して本実施例の原文加工処理について説明する。

【００８０】まず、原文加工部２０は、原文データベース６から文書を構成する各文を読み込む（ステップＳ１１０）。原文データは、例えば図３２（B）に示すよう

15

に、「…首相は、昨夜の首相官邸における…」等の文である。

【００８１】ここで、原文加工部２０の解析部２０ｃは、予め用意された解析辞書２４を参照して既存の言語解析処理を実行し、重文や複文から単文を抽出したり、主語や固有名詞、時間などを表現形態等の構成要素（構成単位）の表現を抽出する（ステップＳ１１１）。

【００８２】型判定部２０ａは型判定用辞書8を参照し、各構成単位に対して型を判定し、その型を示すタグを付与する（ステップＳ１１２）。整形部２０ｂは、まず接続詞辞書9に記述されている接続詞や接続助詞などを文や節から除去し、次に時間を表す型「ｗｈｅｎ」にマッチした文の構成単位の正規化などを行う（ステップＳ１１３）。そして、原文加工部２０は型や言語情報を付与された抜粋文と共に、時間や主語等の構成単位を出力する（ステップＳ１１４）。

【００８３】具体例として、図３０に示すような抜粋情報データが作成されて、抜粋情報データベース２５に格納される。この具体例では、第一文の前半である「日本の政治は、…であるが、」という部分は、「事実」という型に該当し、一方、後半の「この背景には、…という事実がある。」という部分は、「背景」という型に該当している。このようにひとつの文に複数の種類の情報が含まれる場合にも対処できる。また、前半の「日本の政治は、…であるが、」という節末尾には接続助詞や読点がついているが、整形部はこれらを除去して「…である」という終止形にしている。

【００８４】さらに、図３０に示すように、第2文からは、「ｗｈｅｎ」という時間に関する情報と「ｗｈｏ」という動作主体を表現する情報が抽出されている。ここで、原文中の「平成6年7月」という表現が、整形部２０ｂにより「１９９４．７」という形式に正規化されている。

【００８５】本実施例では、図３２（Ｄ）に示すように、型判定用辞書8には例えば5 W1 Hのような抜粋情報の型とそれに対応するパターンが用意されている。この場合、図２８に示すように、品詞情報などの言語情報を用いてパターンを記述してもよい。

【００８６】このような型判定用辞書8を利用して、図３２（Ａ）に示すように、予め用意された5 W1 Hの抜粋情報の型から、利用者が例えば「ｗｈｏ」という動作主体を表現する型が選択されると、原文から例えば「首相」、「前首相」、「科学技術庁長官」等のなんらかの動作の主体となっている人物の名前が抽出されて抜粋情報画面に表示される。同様に、利用者が例えば「ｗｈｅｒｅ」という場所を表現する型が選択されると、同一原文中から例えば「首相官邸」、「国内の原子力発電所」等の場所を表現する言葉が抽出されて抜粋情報画面に表示される。

【００８７】

16

【発明の効果】以上詳述したように本発明によれば、フルテキストの文書を検索する文書検索システムにおいて、第1に、検索された文書の全文から所望の抜粋文や情報のみを抽出して表示できるようにして、利用者の文書検索作業に要する負荷の軽減化を図ることができる。第2に、同一文書に対して様々な観点や基準に基づいて抜粋文や情報を抽出して表示できるようにして、多様な検索目的に応じた検索処理を実現することができる。第３に、検索された文書の全文から所定の基準に基づいて抜粋文や情報のみを抽出して表示する場合に、優先度の順序に従って表示する抜粋文や情報の量を調整できる検索処理を実現することができる。

【００８８】換言すれば、利用者の望む型の文や文の構成要素のみを抜粋して表示することにより、大量の検索結果のブラウジングを効率化することができる。さらに、抜粋情報のみをコンパクトに表示することにより、文書全体を直観的にかつ総括的にとらえることを可能にし、画面のスクロールなどに関する労力及び負荷を軽減することができる。さらに、原文中でどの文が重要かは観点によっても変わるし個人差もあるが、抜粋情報の型の切替により、様々な観点から重要な箇所を抜粋することが可能である。

【図面の簡単な説明】

【図１】本発明の実施例に係わる文書検索システムの要部を示すブロック図。

【図２】第1の実施例に係わる機能ブロック図。

【図３】第1の実施例の動作を説明するための概念図。

【図４】第1の実施例の動作を説明するためのフローチャート。

【図５】第1の実施例の動作を説明するための概念図。

【図６】第1の実施例の動作を説明するための概念図。

【図７】第1の実施例の抜粋処理に関する動作を説明するためのフローチャート。

【図８】第1の実施例のパターン処理に関する動作を説明するための図。

【図９】第1の実施例の選択表示処理に関する動作を説明するためのフローチャート。

【図１０】第1の実施例の選択表示処理に関する動作を説明するための図。

【図１１】第1の実施例の特定文表示処理に関する動作を説明するためのフローチャート。

【図１２】第1の実施例の分量制御に関する動作を説明するための図。

【図１３】第1の実施例の分量制御に関する動作を説明するためのフローチャート。

【図１４】第1の実施例の分量制御に関する動作を説明するための図。

【図１５】第1の実施例の更新処理に関する動作を説明するためのフローチャート。

【図１６】第1の実施例の更新処理に関する動作を説明

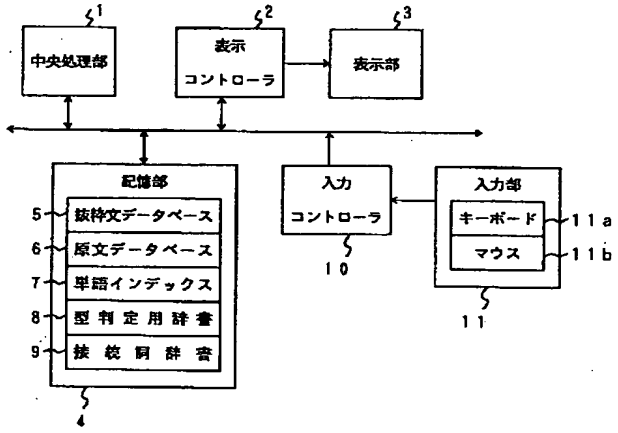
17

するための図。
【 図17 】第1 の実施例の更新処理に関する動作を説明するための図。
【 図18 】第1 の実施例の優先度修正処理に関する動作を説明するためのフローチャート。
【 図19 】第1 の実施例の優先度修正処理に関する動作を説明するための図。
【 図20 】第1 の実施例の複数文書の検索処理に関する動作を説明するためのフローチャート。
【 図21 】第1 の実施例の複数文書の検索処理に関する動作を説明するための図。
【 図22 】第2 の実施例に係わる機能ブロック図。
【 図23 】第2 の実施例の動作を説明するためのフローチャート。
【 図24 】第2 の実施例の動作を説明するための図。
【 図25 】第3 の実施例に係わる機能ブロック図。
【 図26 】第3 の実施例の動作を説明するためのフローチャート。

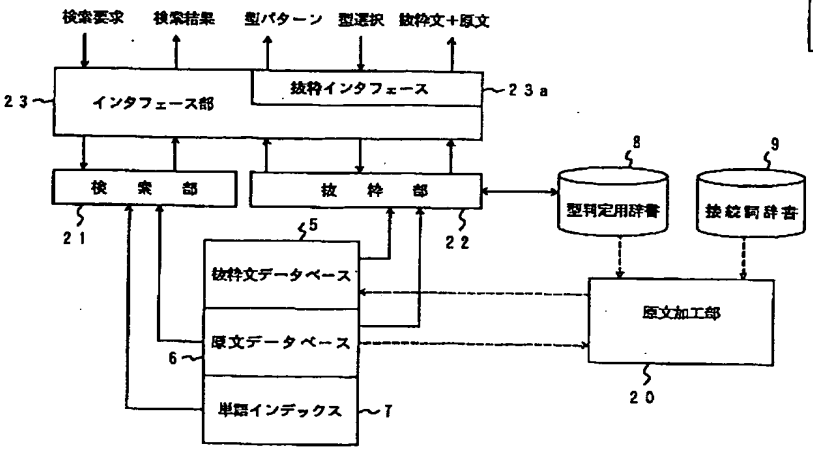
18

【 図27 】第3 の実施例の動作を説明するための図。
【 図28 】第3 の実施例の型判定用辞書の構造を説明するための図。
【 図29 】第4 の実施例に係わる機能ブロック図。
【 図30 】第4 の実施例の動作を説明するための図。
【 図31 】第4 の実施例の動作を説明するためのフローチャート。
【 図32 】第4 の実施例の動作を説明するための図。
【 符号の説明】
1 …中央処理部(ＣＰＵ) 、2 …表示コントローラ、3 …表示部、4 …記憶部、5 …抜粋文データベース、6 …原文データベース、7 …単語インデックス、8 …型判定用辞書、9 …接続司辞書、10 …入力コントローラ、11 …入力部、20 …原文加工部、21 …検索部、22 …抜粋部、23 …インターフェース部、23a …抜粋インターフェース。

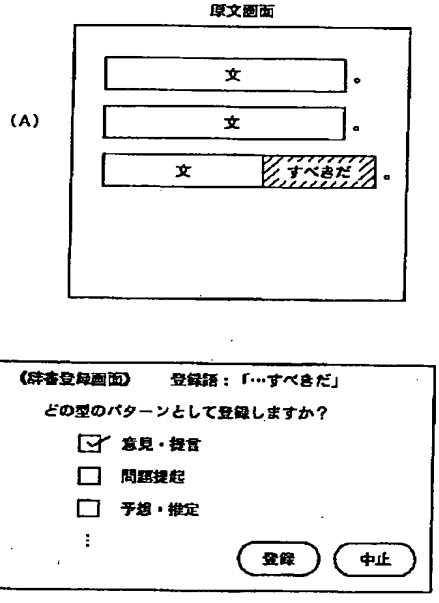
【 図1 】



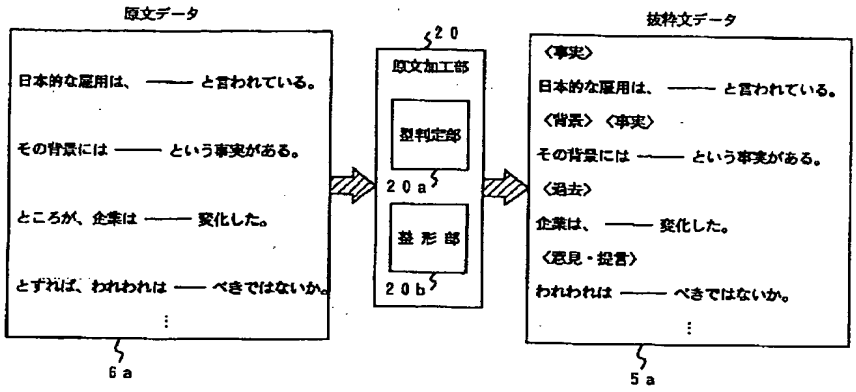
【 図2 】



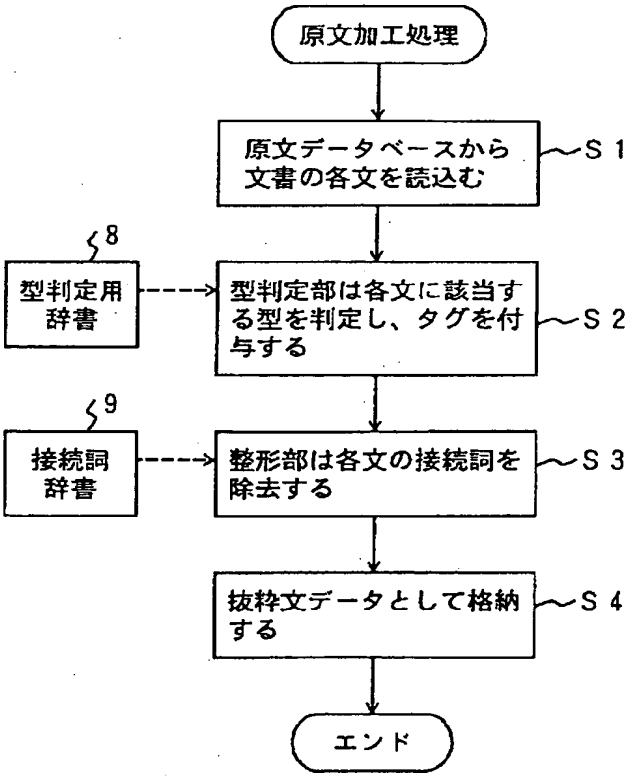
【 図17 】



【図3】



【図4】

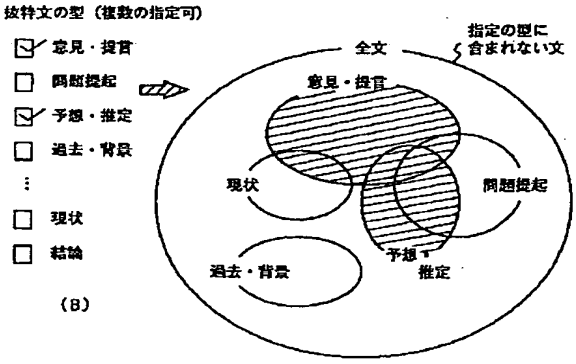


【図5】

(A)

〈型〉	〈パターン〉
意見・提言	「…だと考えられる」 「…べきではないだろうか」 「…を提案する」 「…であると思う」 ⋮
問題提起	「…問題がある」 「…について考えてみたい」 ⋮
予想・推定	「おそらく、…」 「…なるであろう」 「…かも知れない」 「…であると思う」 ⋮
⋮	⋮

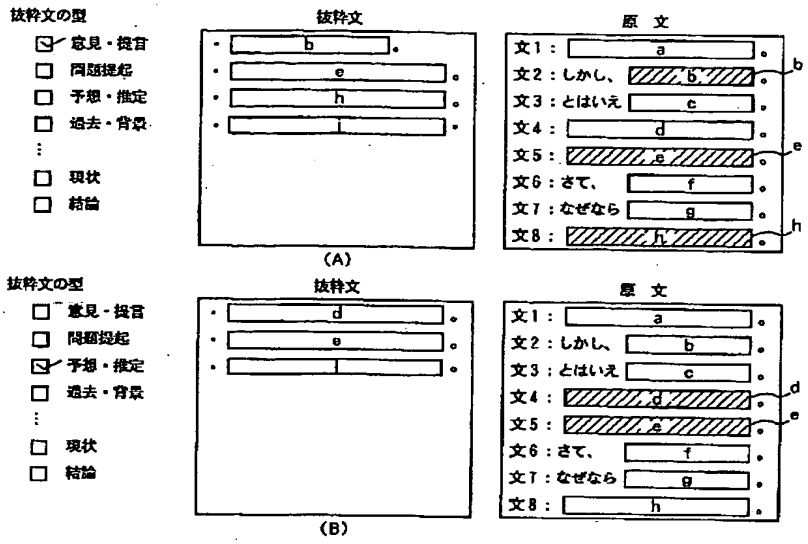
(B)



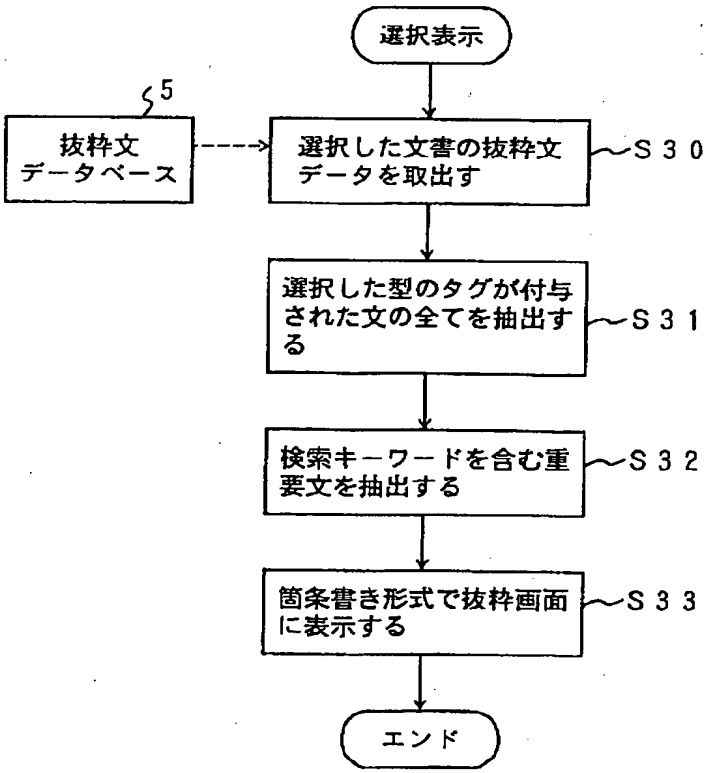
【図28】

〈型〉	〈パターン〉
事実	文末に断定の助動詞がある
予想・推定	文頭に推定の副詞がある 文末に推定の助動詞がある
⋮	⋮

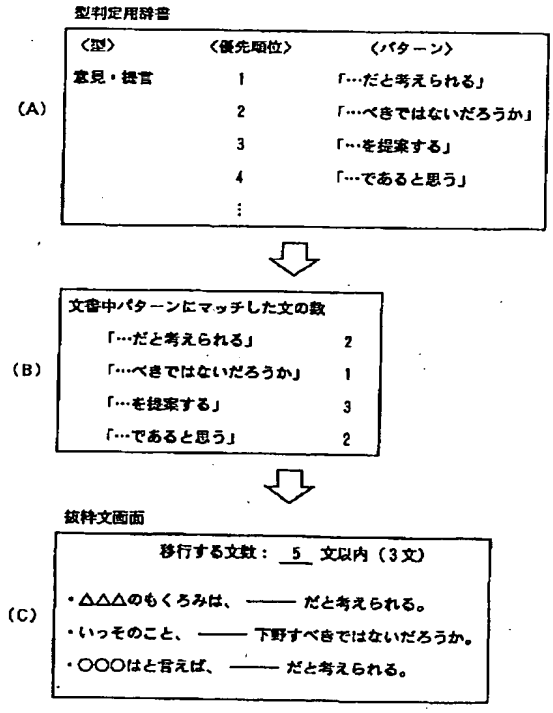
【図6】


抜粋文の型
意見・提言
問題提起
予想・推定
過去・背景
現状
結論
抜粋文
原文
(A)
(B)

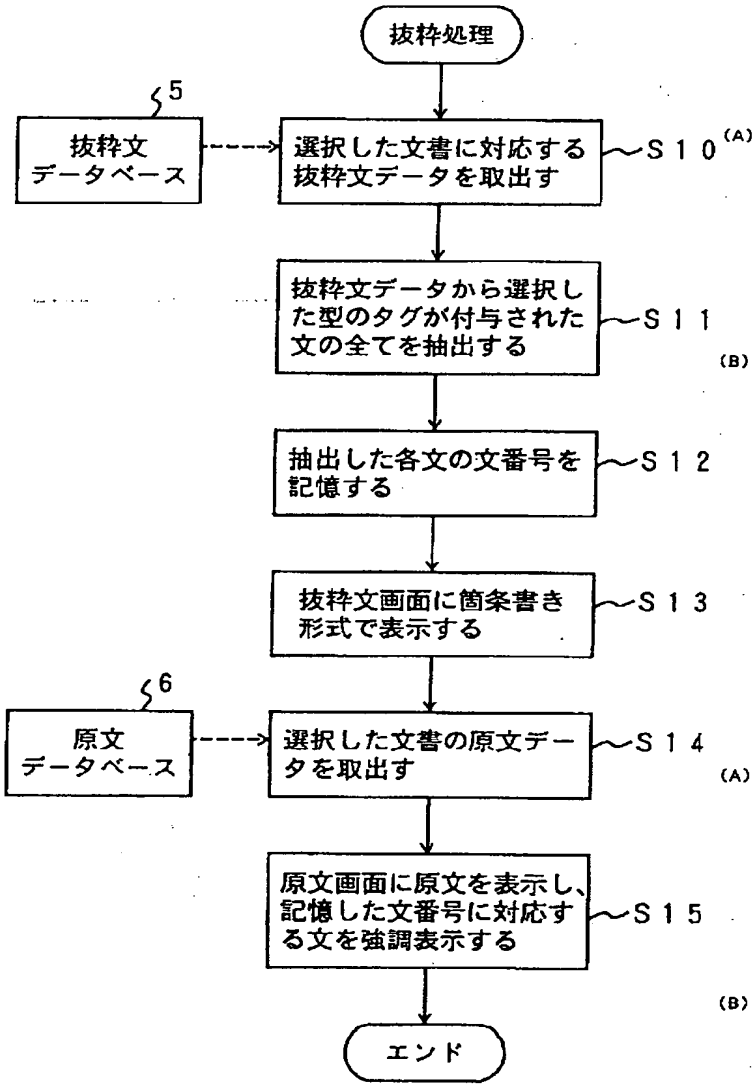
【図9】


選択表示
抜粋文データベース
選択した文書の抜粋文データを取出す
S30
選択した型のタグが付与された文の全てを抽出する
S31
検索キーワードを含む重要文を抽出する
S32
箇条書き形式で抜粋画面に表示する
S33
エンド

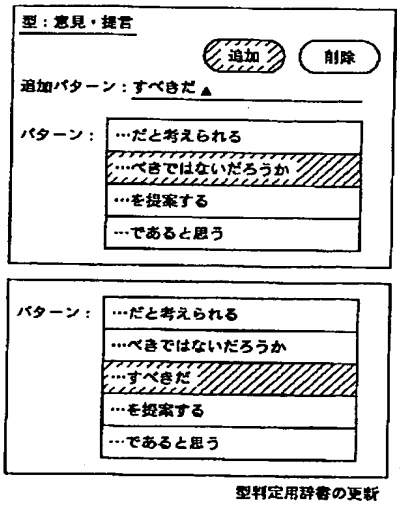
【図14】


型判定用辞書
(A)
(B)
(C)
抜粋文画面

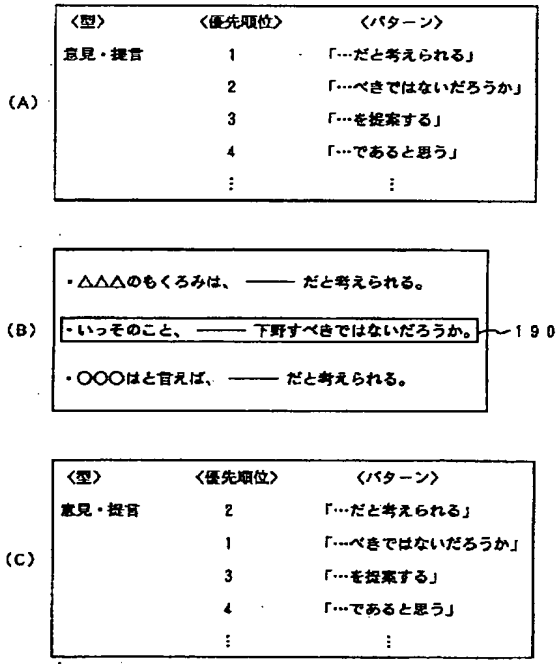
【図7】



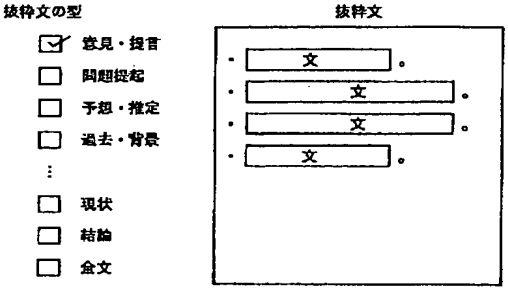
【図16】



【図19】

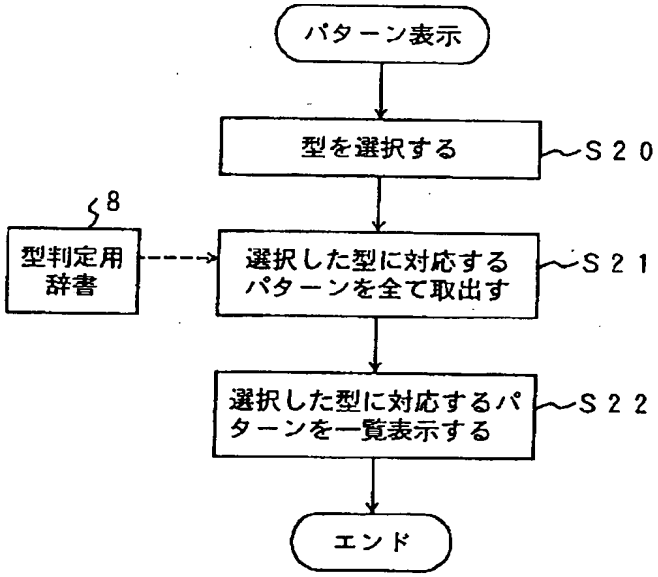


【図24】

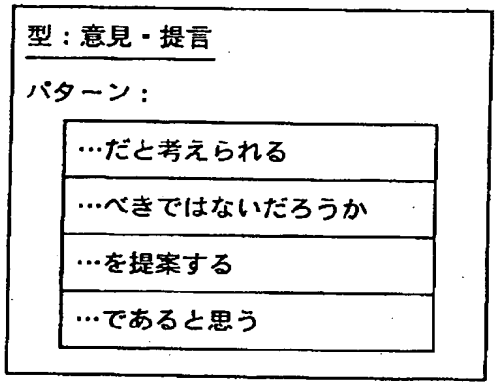


【図8】

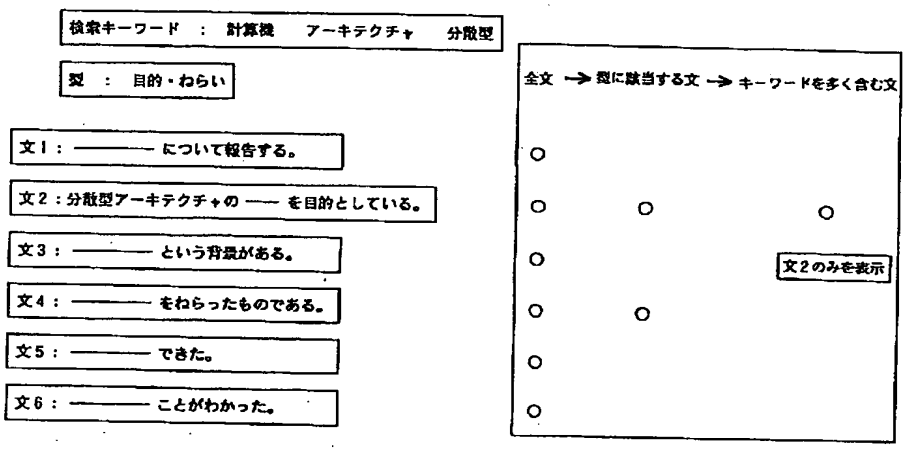
(A)


パターン表示
型を選択する
S２０
8
型判定用
辞書
選択した型に対応する
パターンを全て取出す
S２１
選択した型に対応するパ
ターンを一覧表示する
S２２
エンド

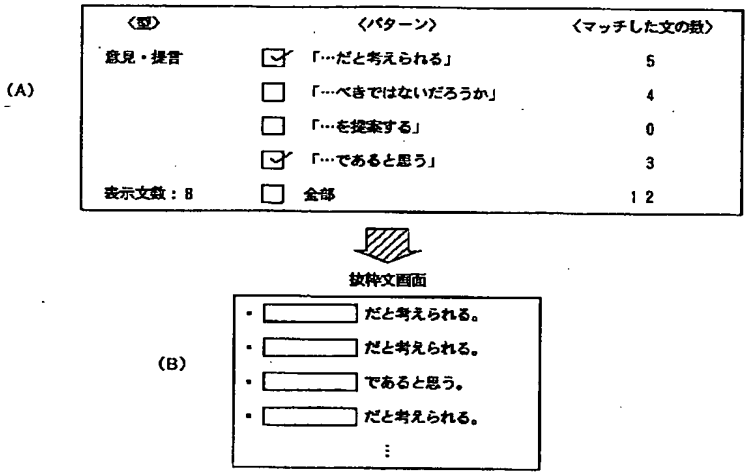
(B)


型：意見・提言
パターン：
…だと考えられる
…べきではないだろうか
…を提案する
…であると思う

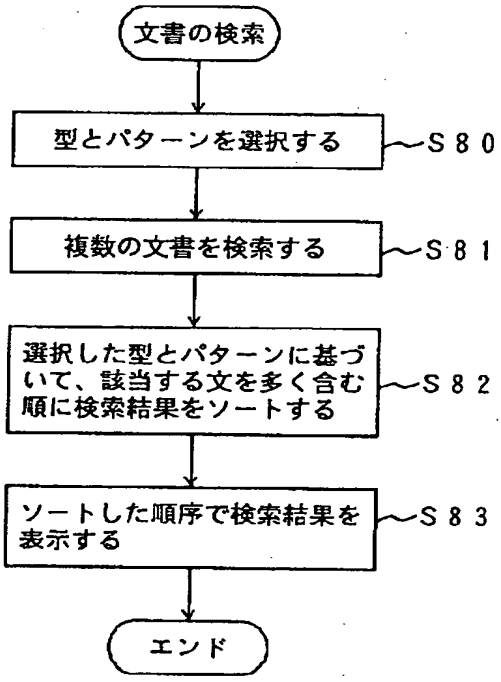
【図１０】


検索キーワード ： 計算機　アーキテクチャ　分散型
型 ： 目的・ねらい
文１：――― について報告する。
文２：分散型アーキテクチャの ―― を目的としている。
文３：――― という背景がある。
文４：――― をねらったものである。
文５：――― できた。
文６：――― ことがわかった。
全文 → 型に該当する文 → キーワードを多く含む文
文２のみを表示

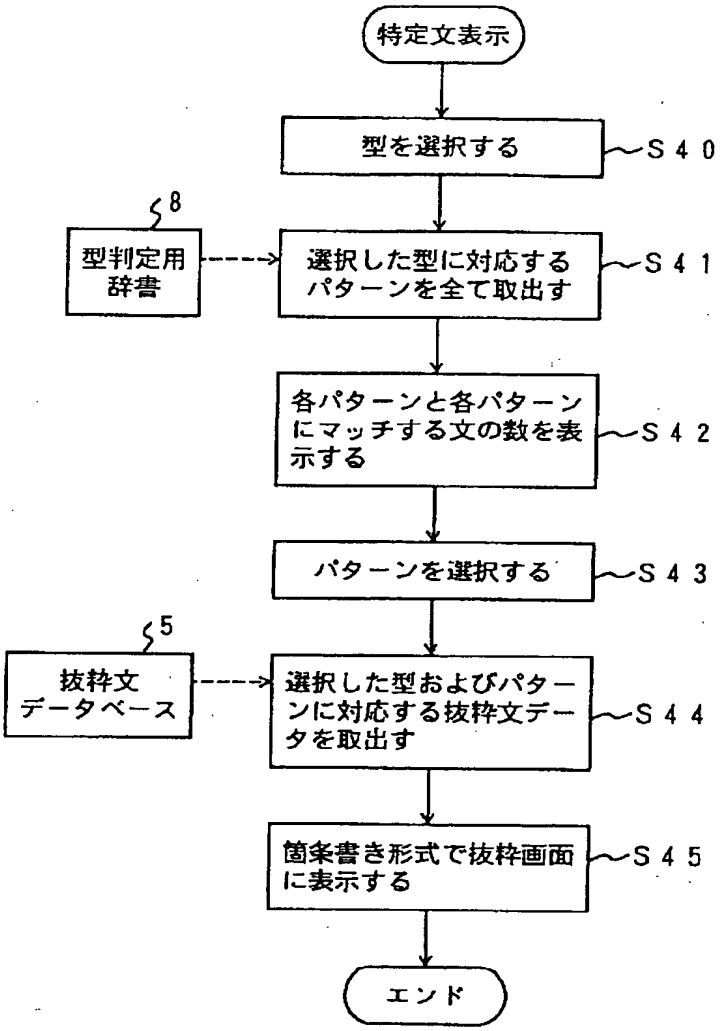
【図１２】


(A)
〈型〉
〈パターン〉
〈マッチした文の数〉
意見・提言
「…だと考えられる」 5
「…べきではないだろうか」 4
「…を提案する」 0
「…であると思う」 3
表示文数：８
全部 12
抜粋文画面
(B)
だと考えられる。
だと考えられる。
であると思う。
だと考えられる。

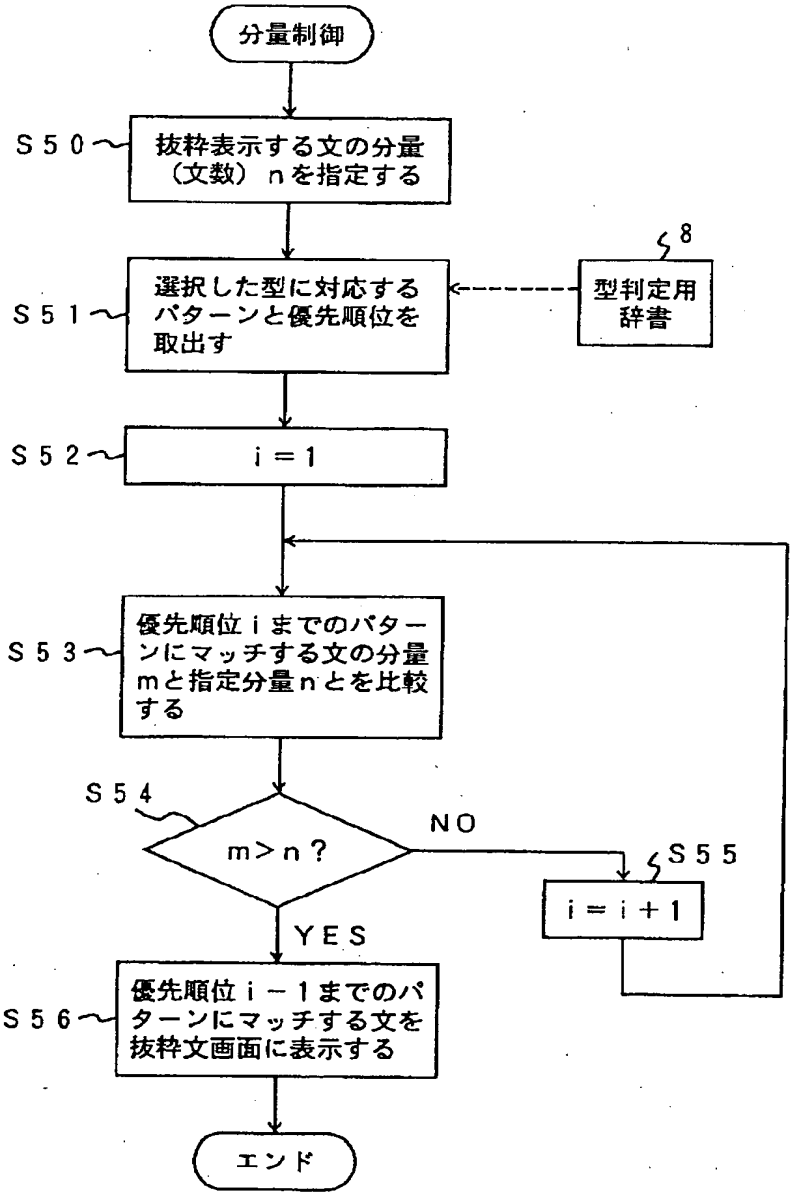
【図２０】


文書の検索
型とパターンを選択する
S８０
複数の文書を検索する
S８１
選択した型とパターンに基づいて、該当する文を多く含む順に検索結果をソートする
S８２
ソートした順序で検索結果を表示する
S８３
エンド

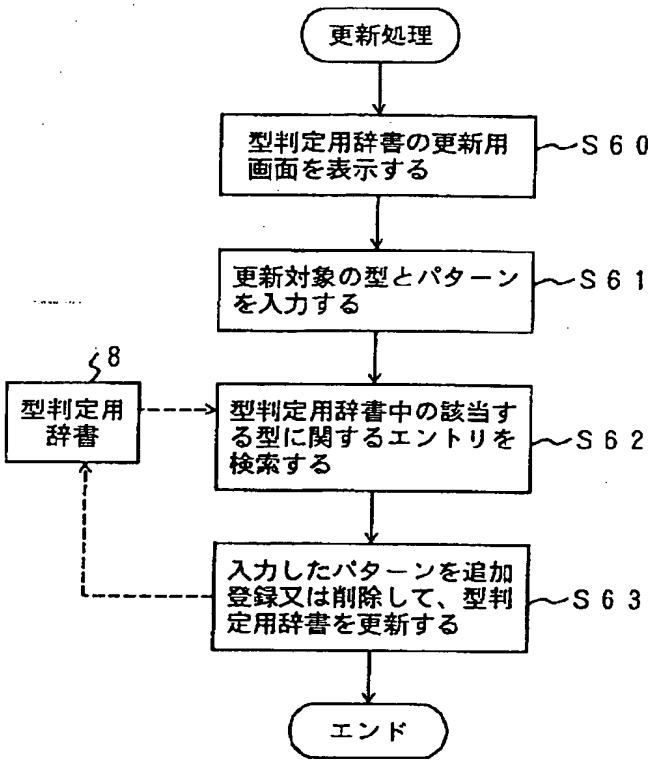
【 図1 1 】


特定文表示
型を選択する
S 4 0
8
型判定用辞書
選択した型に対応するパターンを全て取出す
S 4 1
各パターンと各パターンにマッチする文の数を表示する
S 4 2
パターンを選択する
S 4 3
5
抜粋文データベース
選択した型およびパターンに対応する抜粋文データを取出す
S 4 4
箇条書き形式で抜粋画面に表示する
S 4 5
エンド

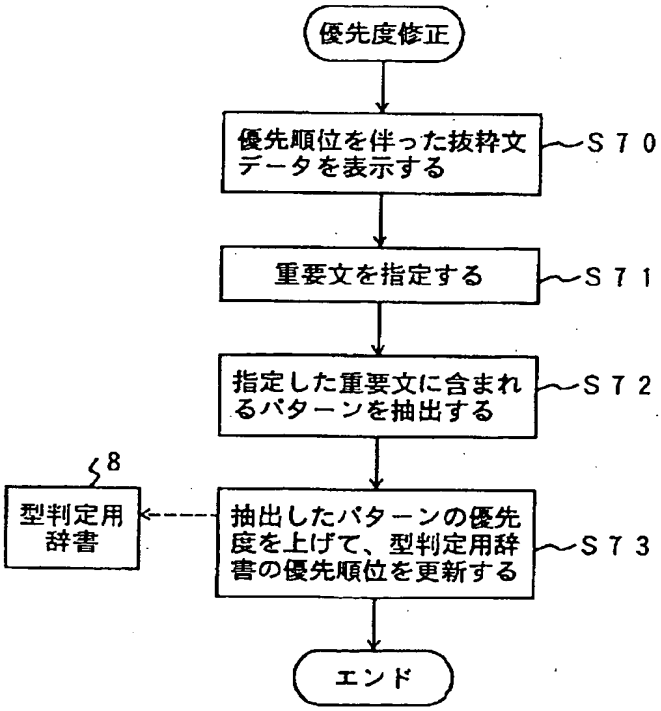
【図13】


分量制御
S50
抜粋表示する文の分量（文数）nを指定する
8
型判定用辞書
S51
選択した型に対応するパターンと優先順位を取出す
S52
i＝1
S53
優先順位iまでのパターンにマッチする文の分量mと指定分量nとを比較する
S54
m＞n？
NO
S55
i＝i＋1
YES
S56
優先順位i－1までのパターンにマッチする文を抜粋文画面に表示する
エンド

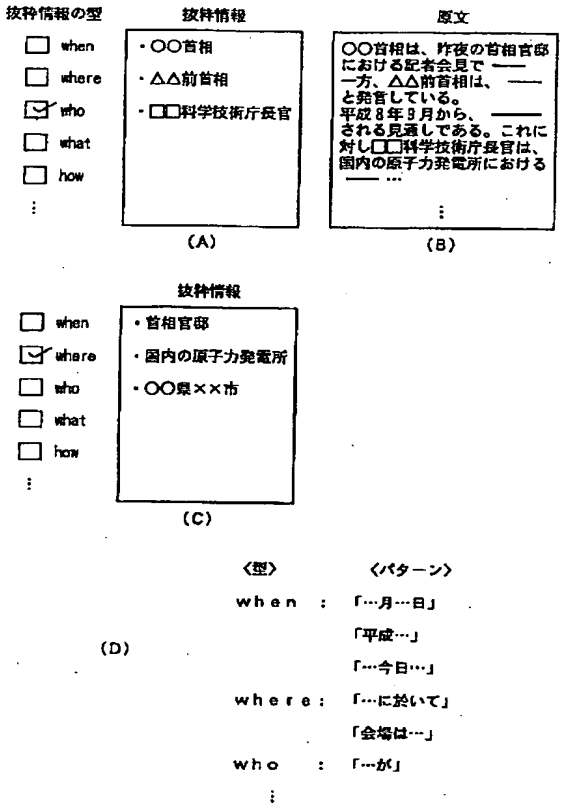
【図１５】


更新処理
型判定用辞書の更新用画面を表示する
S６０
更新対象の型とパターンを入力する
S６１
8
型判定用辞書
型判定用辞書中の該当する型に関するエントリを検索する
S６２
入力したパターンを追加登録又は削除して、型判定用辞書を更新する
S６３
エンド

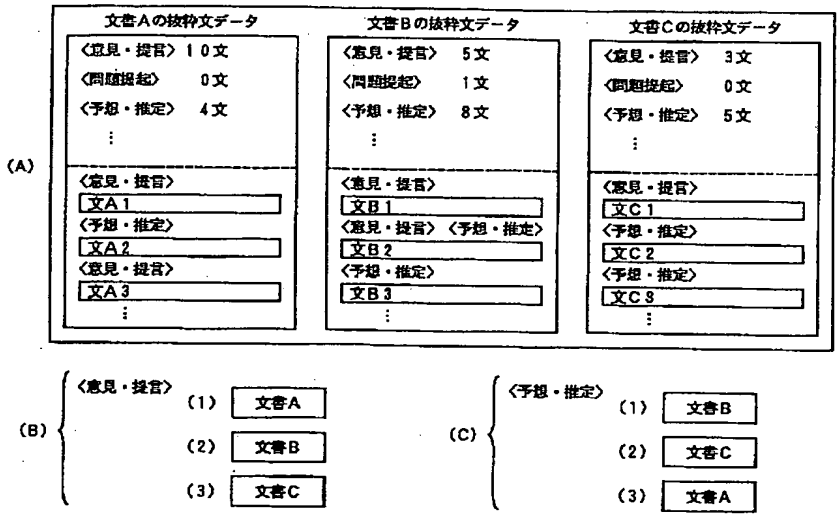
【図１８】


優先度修正
優先順位を伴った抜粋文データを表示する
S７０
重要文を指定する
S７１
指定した重要文に含まれるパターンを抽出する
S７２
8
型判定用辞書
抽出したパターンの優先度を上げて、型判定用辞書の優先順位を更新する
S７３
エンド

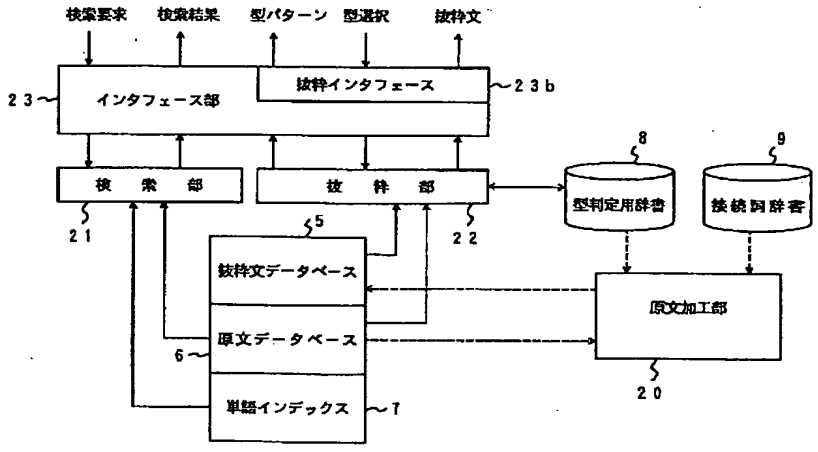
【図３２】


抜粋情報の型
when
where
who
what
how
抜粋情報
・○○首相
・△△前首相
・□□科学技術庁長官
(A)
原文
○○首相は、昨夜の首相官邸における記者会見で ――
一方、△△前首相は、 ――
と発言している。
平成８年９月から、 ――
される見通しである。これに対し□□科学技術庁長官は、国内の原子力発電所における ――
(B)
抜粋情報
・首相官邸
・国内の原子力発電所
・○○県××市
(C)
(D)
〈型〉　〈パターン〉
when : 「…月…日」
「平成…」
「…今日…」
where : 「…に於いて」
「会場は…」
who : 「…が」

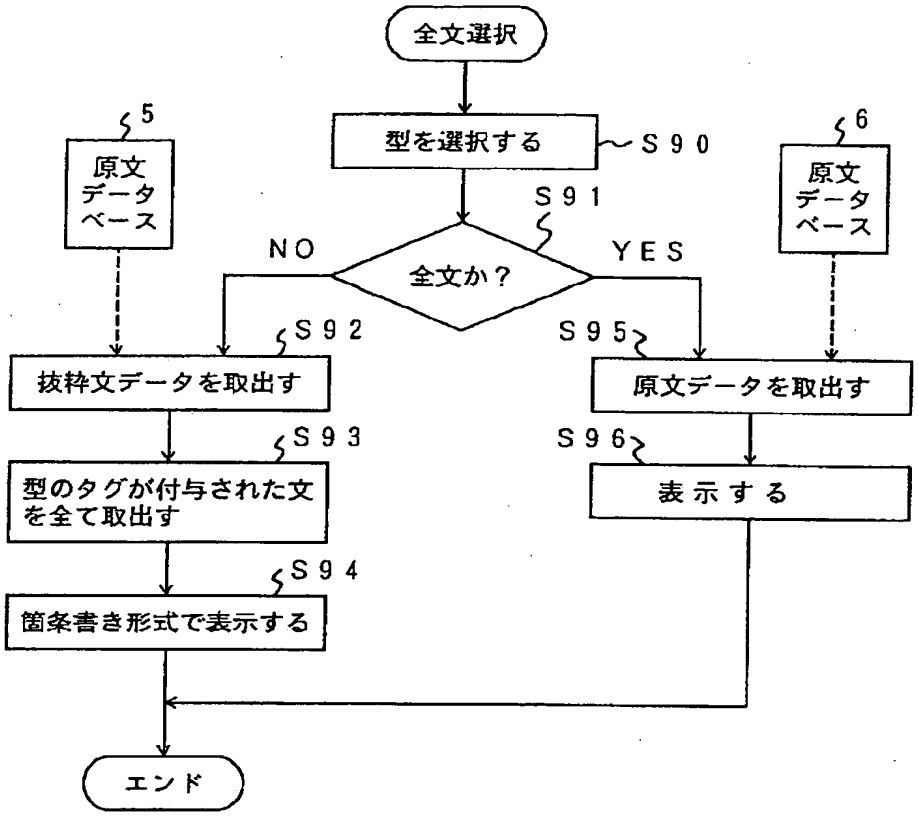
【 図2１ 】


(A)
文書Aの抜粋文データ
〈意見・提言〉 10文
〈問題提起〉 0文
〈予想・推定〉 4文
〈意見・提言〉
文A1
〈予想・推定〉
文A2
〈意見・提言〉
文A3
文書Bの抜粋文データ
〈意見・提言〉 5文
〈問題提起〉 1文
〈予想・推定〉 8文
〈意見・提言〉
文B1
〈意見・提言〉〈予想・推定〉
文B2
〈予想・推定〉
文B3
文書Cの抜粋文データ
〈意見・提言〉 3文
〈問題提起〉 0文
〈予想・推定〉 5文
〈意見・提言〉
文C1
〈予想・推定〉
文C2
〈予想・推定〉
文C3
(B)
〈意見・提言〉
(1) 文書A
(2) 文書B
(3) 文書C
(C)
〈予想・推定〉
(1) 文書B
(2) 文書C
(3) 文書A

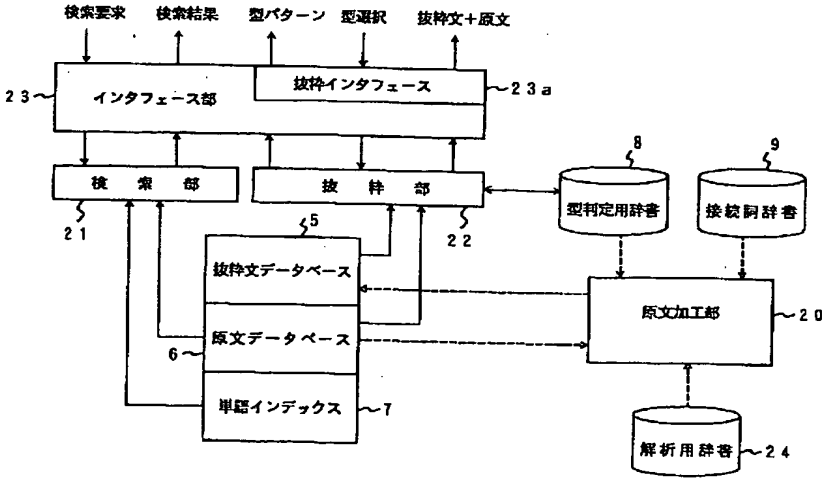
【 図2２ 】


検索要求
検索結果
型パターン
型選択
抜粋文
23
インタフェース部
抜粋インタフェース
23b
検索部
抜粋部
21
22
8
型判定用辞書
9
接続詞辞書
5
抜粋文データベース
原文データベース
6
単語インデックス
7
原文加工部
20

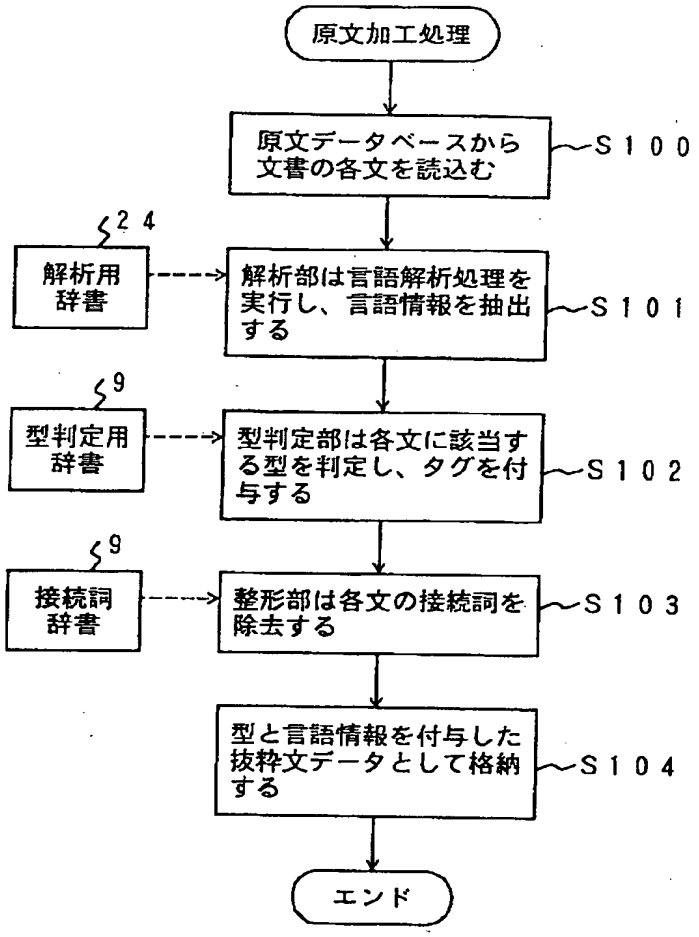
【図２３】


全文選択
型を選択する
S90
S91
全文か？
NO
YES
5
原文データベース
6
原文データベース
S92
抜粋文データを取出す
S93
型のタグが付与された文を全て取出す
S94
箇条書き形式で表示する
S95
原文データを取出す
S96
表示する
エンド

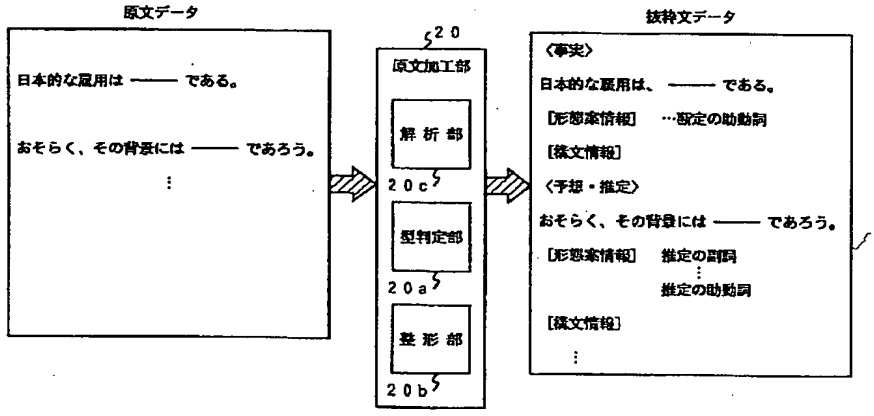
【図２５】


検索要求
検索結果
型パターン
型選択
抜粋文+原文
23
インタフェース部
抜粋インタフェース
23a
検索部
抜粋部
21
22
8
型判定用辞書
9
接続詞辞書
5
抜粋文データベース
原文データベース
6
単語インデックス
7
原文加工部
20
解析用辞書
24

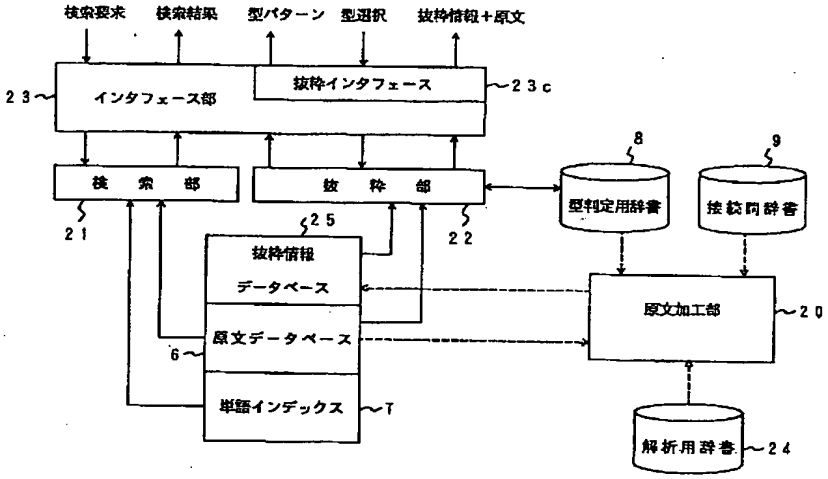
【 図２６ 】


原文加工処理
原文データベースから文書の各文を読込む
S１００
24
解析用辞書
解析部は言語解析処理を実行し、言語情報を抽出する
S１０１
9
型判定用辞書
型判定部は各文に該当する型を判定し、タグを付与する
S１０２
9
接続詞辞書
整形部は各文の接続詞を除去する
S１０３
型と言語情報を付与した抜粋文データとして格納する
S１０４
エンド

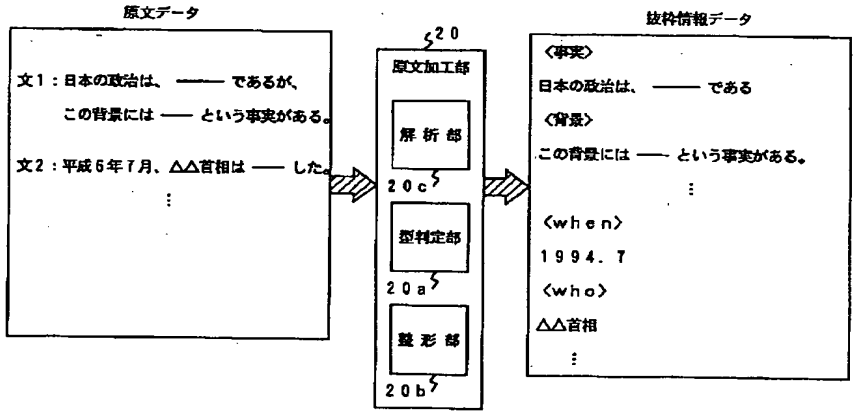
【 図２７ 】


原文データ
日本的な雇用は ―― である。
おそらく、その背景には ―― であろう。
20
原文加工部
解析部
20c
型判定部
20a
整形部
20b
抜粋文データ
〈事実〉
日本的な雇用は、 ―― である。
[形態素情報] …断定の助動詞
[構文情報]
〈予想・推定〉
おそらく、その背景には ―― であろう。
[形態素情報] 推定の副詞
推定の助動詞
[構文情報]

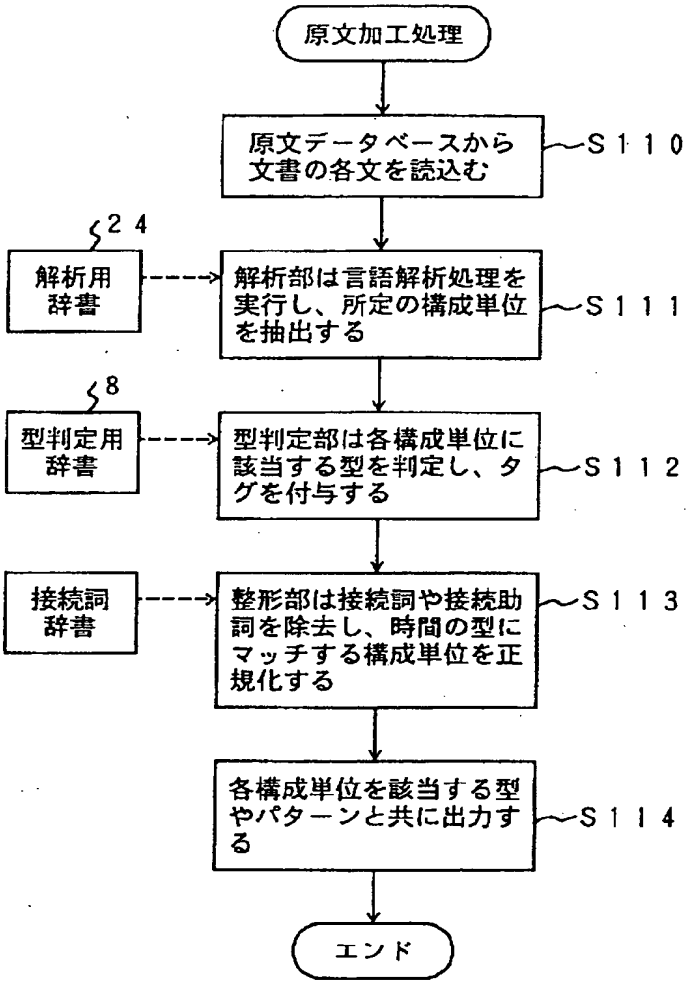
【 図2 9 】


検索要求
検索結果
型パターン
型選択
抜粋情報＋原文
23
インタフェース部
抜粋インタフェース
23c
検　索　部
抜　粋　部
21
22
25
抜粋情報
データベース
原文データベース
6
単語インデックス
7
8
型判定用辞書
9
接続詞辞書
原文加工部
20
解析用辞書
24

【 図3 0 】


原文データ
文1：日本の政治は、 ―― であるが、
この背景には ―― という事実がある。
文2：平成6年7月、△△首相は ―― した。
20
原文加工部
解析部
20c
型判定部
20a
整形部
20b
抜粋情報データ
〈事実〉
日本の政治は、 ―― である
〈背景〉
この背景には ―― という事実がある。
〈when〉
1994. 7
〈who〉
△△首相

【 図3 1 】


原文加工処理
原文データベースから文書の各文を読込む
S 1 1 0
24
解析用辞書
解析部は言語解析処理を実行し、所定の構成単位を抽出する
S 1 1 1
8
型判定用辞書
型判定部は各構成単位に該当する型を判定し、タグを付与する
S 1 1 2
接続詞辞書
整形部は接続詞や接続助詞を除去し、時間の型にマッチする構成単位を正規化する
S 1 1 3
各構成単位を該当する型やパターンと共に出力する
S 1 1 4
エンド